Complete evaluation metrics guide for image classification — all metrics, code placement, and monitor names

BLOCK 7 — get_metrics() ফাংশনে বসাবেন + BLOCK 8 callbacks monitor বদলাবেন

Metric	কখন ব্যবহার করবেন	BLOCK 7 → return [এই code] দিন	BLOCK 8 monitor
Accuracy	Balanced Dog Breed, Cassava, Plant — সব class-এ ডেটা সমান	'accuracy'	val_accuracy (default)
Top-K Accuracy	100+ classes Dog Breed (120 class) — প্রথম ৫টায় ১টা মিললেই হয়। Template auto-যোগ করে।	✓ template auto-adds TopKCategoricalAccuracy(k=5, name='top5_acc')	val_top5_acc
AUC-ROC	Imbalanced Skin Cancer, Fraud — ৯৯% normal ১% রোগ। Accuracy দিলে চিট হবে!	tf.keras.metrics.AUC(name='auc')	val_auc
AUC Multilabel	Multilabel এক ছবিতে একাধিক tag — প্রতিটি label আলাদাভাবে AUC মাপে	tf.keras.metrics.AUC(multi_label=True, name='auc')	val_auc
Binary Accuracy	Multilabel Multilabel task-এ sigmoid output এর জন্য	tf.keras.metrics.BinaryAccuracy(name='accuracy')	val_accuracy
F1-Score (Macro)	Hackathon Rules-এ "F1 macro" লেখা থাকলে এটা দিন	tf.keras.metrics.F1Score(average='macro', name='f1_score') ⚠ TF 2.16+ only	val_f1_score
Precision	Medical False positive কমাতে — ভুল করে	tf.keras.metrics.Precision(name='precision'	val_precision

	রোগ ধরা যাবে না	) ⚠ Binary default — multiclass-এ Section B দেখুন	
Recall (Sensitivity)	Medical False negative কমাতে — রোগী miss করা একদম চলবে না	tf.keras.metrics.Recall(name='recall') ⚠ Binary default — multiclass-এ Section B দেখুন	val_recall

_evaluate_model() ফাংশনের ভেতরে — training শেষ হলে sklearn দিয়ে compute হয়

Metric	কখন ব্যবহার করবেন	_evaluate_model()-এ y_pred, y_true এর পরে যোগ করুন	উপযুক্ত competition
Log Loss	Probability sub Submission-এ class নয়, probability দিতে হলে	✓ template-এ আছে log_loss(y_true_oh, y_pred_prob)	Dog Breed State Farm
Classification Report	সব multiclass প্রতিটি class-এর F1, Precision, Recall একসাথে দেখায়	✓ template-এ আছে classification_report(y_true, y_pred, target_names=class_names)	সব competition
Macro AUC (multiclass)	Imbalanced multiclass Keras AUC এর চেয়ে সঠিক। OvR পদ্ধতিতে প্রতিটি class আলাদা।	from sklearn.metrics import roc_auc_score auc = roc_auc_score(y_true_oh, y_pred_prob, multi_class='ovr', average='macro') print(f'Macro AUC: {auc:.4f}')	ISIC Skin Medical
Cohen's Kappa (Quadratic)	Ordinal class ০→১→২→৩→৪ severity — ভুলের দূরত্ব গুরুত্বপূর্ণ। Off-by-1 আর off-by-4 আলাদা।	from sklearn.metrics import cohen_kappa_score kappa = cohen_kappa_score(y_true, y_pred, weights='quadratic') print(f'QWK: {kappa:.4f}')	APTOS Eye Prostate cancer
MCC (Matthews)	Severely imbalanced AUC এর চেয়ে কঠিন। সব ধরনের error ধরতে পারে। -1 থেকে +1।	from sklearn.metrics import matthews_corrcoef mcc = matthews_corrcoef(y_true, y_pred	ISIC Skin Fraud detection

		```) print(f'MCC: {mcc:.4f}') ```	

mAP
(Mean Avg Precision)
	Multilabel/Ranking
	প্রতিটি class-এর Average Precision গড় করে। mAP=1 মানে perfect।
		``
		from sklearn.metrics import average_precision_score
		map_s = average_precision_score(
		  y_true_oh, y_pred_prob,
		  average='macro'
		)
		print(f'mAP: {map_s:.4f}')
		``
			Multilabel tag
			Object ranking

BLOCK 1 — LOSS_BASE = এই code দিন (training এর direction বদলায়)

Loss function	কখন ব্যবহার করবেন	BLOCK 1-এ LOSS_BASE = এই code	Note
Categorical Cross-Entropy	Default ৯৯% competition — standard multiclass। Label smoothing overfitting কমায়।	✓ template default ``` CategoricalCrossentropy(label_smoothing=0.1) ```	পরিবর্তন নেই
Binary Cross-Entropy	Multilabel task_type='multilabel' দিলে template auto-set করে	✓ template auto-set ``` BinaryCrossentropy(label_smoothing=0.1) ```	পরিবর্তন নেই

Focal Loss	Imbalanced
	class_weight যথেষ্ট না হলে। Hard example-এ বেশি focus করে।
		``
		def focal_loss(gamma=2.0, alpha=0.25):
		  def fn(y_true, y_pred):
		  y_pred=tf.clip_by_value(y_pred,1e-7,1-1e-7)
		  ce=-y_true*tf.math.log(y_pred)
		  w=alpha*y_true*tf.pow(1-y_pred,gamma)
		  return tf.reduce_mean(tf.reduce_sum(w*ce,axis=-1))
		  return fn
		LOSS_BASE = focal_loss(gamma=2.0)
		``
			Skin Cancer
			Fraud
			gamma↑ = harder examples-এ বেশি focus